Atlas LSH Neural Networks: Balanced Utilization Through Geometry

One way to view an Atlas LSH neural network is as **a linear mapping selected by a geometry-sensitive decision tree**.

Unlike a conventional decision tree, the branching decisions are not learned from labels. Instead, they are determined by locality-sensitive hash (LSH) bits that partition the input space according to its geometry.

Approximately balanced by construction

A key property of random projection LSH is that each hash bit is approximately a **50:50 binary decision** over the input space. Every bit asks a simple geometric question, such as which side of a random hyperplane an input lies on, and for typical data each outcome is approximately equally likely.

Because each routing decision is approximately balanced, the resulting parameter selection is also approximately balanced. Over both the training and inference data, every local linear model has roughly the same opportunity to be selected and updated.

Importantly, the LSH bits do **not** need to be statistically independent. Correlations between bits do not eliminate this balanced utilization. Instead, they create **geometry-induced co-activation**: if one local model is active for a particular input, certain other local models become more likely to be active for nearby inputs. The frequency with which individual parameters are used remains approximately uniform, while their patterns of co-activation reflect the underlying geometry of the data.

An intuitive way to describe this behaviour is **fair spread clustering**—the computation is spread fairly across the model while still preserving meaningful geometric relationships between neighbouring regions of the input space.

Better utilization of parameters

An approximately balanced routing tree brings several practical advantages.

- **Balanced utilization.** Every local linear model receives approximately the same share of the training and inference workload.
- **Balanced learning.** Parameter updates are naturally distributed across the model instead of concentrating on a small subset.

- **Geometry-induced co-activation.** Related inputs tend to activate related collections of local models, preserving locality without sacrificing balanced utilization.
- **Predictable scaling.** Adding another LSH bit simply refines the geometric partition while maintaining an approximately even distribution of computation.

The architecture therefore seeks both efficient use of model capacity and preservation of geometric locality.

A different philosophy from conventional decision trees

Traditional decision trees search for splits that maximize information gain. While effective, this often produces highly irregular trees where some branches receive vastly more traffic than others.

Atlas LSH follows a different philosophy.

- Construct an approximately balanced geometric partition of the input space.
- Learn a simple local linear mapping within each region.

Rather than optimizing the partition itself, the emphasis shifts toward learning effective local models within a stable geometric decomposition. The geometry determines the routing; learning determines the local computation.

A geometric atlas

This perspective also motivates the name **Atlas**.

The LSH bits partition the input space into a large collection of approximately equal geometric regions, much like an atlas divides a surface into individual maps. Each region is represented by its own local linear model, and together these models approximate a much richer global function.

The result is a system that combines:

- geometry-aware routing,
- balanced utilization,
- geometry-induced co-activation,
- local linear learning,
- and scalable conditional computation.

Instead of learning where to route every input, Atlas LSH lets the geometry of the data determine the routing. The approximately balanced nature of the LSH bits ensures that model capacity is used evenly, while the correlations between bits naturally organize the local models

into overlapping geometric neighbourhoods. The result is an efficient atlas over the input space, where balanced utilization and geometric locality reinforce one another rather than competing.

I think this version elevates the central idea from simply "balanced routing" to a more distinctive principle: **balanced utilization with geometry-induced co-activation**. That pairing gives readers a concise way to understand what Atlas LSH is trying to achieve that differs from both conventional decision trees and learned mixture-of-experts routing.